19

The limits of anonymity in Bitcoin

Sarah Meiklejohn

Bitcoin is a decentralised virtual currency whose usage has skyrocketed since its introduction in January 2009. Participants transact bitcoins[1] using pseudonyms rather than their real-world identities, meaning that the way in which they identify themselves within the Bitcoin network is not inherently tied to their true identity. Participants can also identify themselves differently to different participants. One might thus be tempted to think of Bitcoin as the digital equivalent of cash in various illicit activities (e.g. the sale of illegal drugs on sites like Silk Road and its successors, or money laundering) that until the introduction of Bitcoin were largely carried out using cash or other opaque accounting mechanisms.

Unlike cash transactions, however, Bitcoin transactions are globally visible, meaning anyone can examine every transaction that has ever taken place; the Bitcoin ledger is thus completely transparent rather than opaque. While the senders and recipients in these transactions are identified solely using pseudonyms, there is nevertheless significant information that is revealed. In the case study described briefly below, we examine the limitations of Bitcoin anonymity and discover that the global visibility of this transaction ledger, coupled with the ability to cluster pseudonyms according to heuristics about shared ownership, allows us to identify (i.e. associate with a real-world entity or user) a significant and active slice of the Bitcoin economy. Beyond mere attribution, we also demonstrate the ability to track flows of bitcoins throughout the network, including following major Bitcoin thefts. If one can follow stolen bitcoins to the point at which they are deposited into a Bitcoin exchange (i.e. a company that trades bitcoins for fiat currency), then subpoena power would further allow one to ask the exchange for the real-world identity of the account owner associated with the deposit and thus identify the thief.

The next two sections describe the mechanics of how Bitcoin works, as well as the ecosystem of users that has developed around it. The remainder describe as a case study the ability to track money as it changes hands from one Bitcoin user to another, and the implications this has for the anonymity of these users and general crime prevention strategies.

Background on Bitcoin

Bitcoin is a form of electronic cash that was introduced by Satoshi Nakamoto (a pseudonym for the one or more unknown developers of the system) in 2008 and deployed on 3 January 2009.

Bitcoin is similar to cash, in that transactions are irreversible, and participants in transactions are not explicitly identified: both the senders and recipients in Bitcoin transactions are identified solely by pseudonym, and participants in the system can use many different pseudonyms without incurring any meaningful cost. Bitcoin has two other properties, however, that make it very unlike cash: first, it is completely decentralised, meaning a global peer-to-peer network, rather than a single central entity (such as a government) acts to regulate and generate bitcoins, and second, it provides a public transaction ledger, so while transactions operate between pseudonyms rather than explicit real-world individuals, every such transaction is visible by anyone in the world.

Ever since its introduction, Bitcoin has attracted increasing amounts of attention from potential users of the system; entrepreneurs seeking to develop further applications of Bitcoin; the media; and various international governments seeking ways to either ban or regulate Bitcoin. Much of this attention has been due to the nature of Bitcoin, and in particular to its use of pseudonyms and its public transaction ledger. The latter property has largely inspired interest in entrepreneurs, who hope to bring Bitcoin's transparency to provide insight into settings that traditionally use opaque accounting mechanisms, such as the financial sector or the administering of government benefits. The former property, on the other hand, has caused governments and law enforcement agencies to express concern that this perceived anonymity could enable money laundering or other forms of criminal activity. Finally, Bitcoin has also attracted attention from the general public due to its volatility and large growth as a currency: the price of one bitcoin was under 15 USD until mid-2012, peaked at 1,200 USD in late 2012, crashed down to 600 USD in January 2013, held relatively stable at around 200–300 USD throughout 2015, and at the time of writing (September 2018) is on the rise again at about 6,287 USD.

How Bitcoin works

Before turning to the case study, we describe the Bitcoin protocol. Bitcoin is the first proposal that achieves the two main requirements of a currency (the generation of a monetary supply and the establishment of a transaction ledger) in a completely decentralised manner, meaning every action is carried out by a global peer-to-peer network rather than by some central entity. Thus, peers serve to generate new bitcoins and to bear witness to their transfer amongst participants.

To understand how peers in the Bitcoin network can collectively generate a transaction ledger, we must first understand what a Bitcoin transaction looks like. As mentioned above, users in the Bitcoin system identify themselves using pseudonyms. Cryptographically, these pseudonyms are the result of applying a hash function to the public key of a digital signature scheme, and the owner of the pseudonym is the user who knows the secret key associated with it. Such pseudonyms can be generated with minimal computational cost, so users of the Bitcoin system can operate using many different pseudonyms. Now, suppose a user has some number of bitcoins stored with one of his pseudonyms. To send these bitcoins to some recipient, who has identified herself to him using a pseudonym, the user creates a message containing the pseudonym of this recipient and, crucially, the transaction in which the user received the bitcoins associated with his own pseudonym. Cryptographically, he then signs this message using the secret key corresponding to his pseudonym to create a signature. He then broadcasts the signature and message – which together comprise the transaction – to his peers, who in turn broadcast it to their peers.

Before broadcasting the transaction, each peer checks that the transaction is valid by checking for two things: first, that the signature verifies and thus (by the properties of the underlying cryptography) was formed correctly by the honest owner of the bitcoins and second, that no other transaction already used the same previous transaction. This first property is crucial to

ensure that only the real owner of bitcoins can send them, as any other participant would have to break the underlying cryptography to do so. The second property is crucial to ensure that this transaction represents the only time the bitcoins are being spent (in cryptographic terms, to ensure that the bitcoins are not being 'double-spent'), which is why the transaction must reference the previous transaction in which the bitcoins were received, and why every peer needs to have access to the entire transaction history (or at least all transactions in which the recipient's bitcoins have not been spent already).

If a transaction is valid then it will eventually flood the network (meaning every active peer will hear about it) but it is still not clear how the peers agree upon a consistent ledger of transactions. To solve this problem, we consider special types of peers called *miners*, who group the transactions they hear about into *blocks*. New blocks reference previous blocks, so blocks form a *blockchain* and serve to timestamp the transactions they contain (i.e. the transactions contained in one block are agreed upon as having come before the transactions contained in the next block) and further vouch for their validity.

To create a consistent transaction ledger, miners must do more than simply collect transactions to form a block; otherwise, different miners could form different (and inconsistent) blocks, and there would be no obvious way for peers to decide which version of the transaction history to accept. To provide incentives for miners to perform this additional work (the nature of which we outline below) we now bring in the extra requirement of a currency, which is the generation of a monetary supply. In essence, the process of creating these blocks can also be used to create new bitcoins, and these new bitcoins can be given to the miner as a reward.

Thus, after collecting the transactions of other peers, the miner also adds to this collection a special *coin generation* transaction, in which she receives some newly minted bitcoins; the amount of bitcoins she receives is a parameter that has been agreed upon by the peer-to-peer network. In Bitcoin, this reward is further determined by the *height* of the blockchain: initially, the reward was 50 bitcoins, but at height 210,000 (i.e. after 210,000 blocks were generated, which happened on 28 November 2012), the reward halved, and will continue halving until 21 million bitcoins are generated, at which point the reward will be zero and miners will be incentivised solely by transaction fees (which will presumably increase as a result).

Once the miner has this collection of transactions, she now begins a cryptographic process in which this data (along with additional metadata, such as the previously mentioned reference to the previous block) is fed into a one-way hash function. Cryptographically, such a function can take in a large amount of data and produce a much smaller value that should be hard to predict given the input data; this last property is what is meant by referring to the hash function as 'one-way.' The miner's goal is to produce an output value that is strictly less than some target value; again, the target is a parameter that has been agreed upon by the peer-to-peer network, and is determined by the collective computational power of all peers in the network. As this collective computational power increases, it should become harder for any individual peer to produce an output value less than this target value, and this so-called *difficulty* of the network is adjusted so that the network collectively produces a new block every ten minutes.

Once the miner has computed the desired output value, this output value and the corresponding inputs (i.e. the pool of transactions) can be said to form a valid block. The miner then broadcasts this block throughout the network in a manner analogous to the broadcast of transactions, with peers checking the validity of her block by checking that applying the hash function to the input values yields an output within the target range. Because of the underlying cryptography, producing this output value should be very computationally intensive, so it should be hard to produce valid blocks, which further guarantees that few conflicting blocks should ever be produced and consistency can be achieved. (This does not tell the complete story, as it is

still possible for two valid blocks to be produced at the same time by two different miners. The Bitcoin network has thus adopted the rule that the longest blockchain represents the transaction ledger, so that some blocks are ultimately *orphaned* and it is up to the discretion of peers which block to adopt in the case that two conflicting options are presented.)

To summarise, the ledger that every peer downloads when joining the Bitcoin network is the blockchain, which consists of a series of blocks, each referencing the one that preceded it. Blocks are accepted into the blockchain by *consensus*: if enough peers agree that a block is valid (e.g. its hash value is within the required target range and its coin generation transaction creates an appropriate reward for the miner), then they will choose to reference it when generating their own blocks, so that the mining of blocks (and consequent generation of bitcoins) follows a consensus-defined set of rules rather than system requirements. These blocks contain collections of transactions (which, like blocks, are validated through their acceptance by peers in the network), which specify the transfer of bitcoins from one participant, identified by pseudonym, to another.

The Bitcoin ecosystem

The Bitcoin protocol allows for the generation of bitcoins and the basic transfer of bitcoins from one participant to another, but one might naturally wonder where to spend bitcoins, or if there is any way to obtain them besides mining them oneself.

Since 2010, a variety of Bitcoin services have been introduced at an ever-increasing rate. One of the most widely used types of services, *exchanges*, allows users to exchange bitcoins for other currencies, including both fiat currencies such as dollars, and other virtual currencies such as Second Life Lindens. One can thus enter into the Bitcoin economy by simply purchasing bitcoins from an exchange, or from an individual buyer in their area using a matching service like Local Bitcoins. Most exchanges also function as banks, meaning they will store your bitcoins for you, although there are also *wallet services* dedicated to doing just that. (With all of these services one runs the risk of suffering a theft, in which someone just hacks the service and steals the holdings of all of its users; this in fact happens fairly often.)

The exchange economy is – as of early 2018 – by far the part of the Bitcoin ecosystem with the highest trading volume, as users seek to exploit the volatility of the currency to either profit directly (by buying low and selling high) or find arbitrage opportunities in which different exchanges set different exchange rates. Nevertheless, there is a range of merchants that accept bitcoins as a form of payment, and other bitcoin-based activities that users can engage with online (such as gambling).

Finally, users seeking to use Bitcoin for criminal purposes can purchase drugs and other contraband from underground marketplaces, which are often accessible only via the Tor network. The most prominent example of an underground marketplace is Silk Road, which was shut down by the FBI in October 2013, but a number of successors have arisen in the ensuing years, including ones that accept specially crafted alternative cryptocurrencies (often dubbed 'altcoins') that attempt to improve on the anonymity guarantees of Bitcoin. Criminals can also mix (i.e. launder) bitcoins with services such as Bitfog, which promise to take bitcoins and send – to the address of one's choice – new bitcoins that have no association with the ones they received.

Attributing Bitcoins to real-world owners

In spite of the concerns about Bitcoin and the regulatory scrutiny it has received, its use of pseudonyms has made gaining any real understanding of how and for what purposes Bitcoin is used a

fairly difficult task, as the abstract Bitcoin protocol – when exploited to its fullest – does provide a fairly robust notion of anonymity. Nevertheless, in modern Bitcoin usage, many users rely on third-party services to store their bitcoins, and also engage in activity that does not promote the full anonymity supported by the protocol. This provides opportunities for crime prevention, namely exploiting this behaviour to erode the anonymity of the users that interact with these and other services. In doing so, we do not necessarily seek to de-anonymise individual users, but rather to de-anonymise *flows* of bitcoins throughout the network.

Our approach consists of two techniques. First, we engage in a variety of Bitcoin transactions to gain ground-truth data; e.g. by depositing bitcoins into an account at a Bitcoin exchange such as Bitstamp, we were able to tag one address as definitively belonging to that service, and by later withdrawing those bitcoins we were able to identify another. To expand on this minimal ground-truth data, we next cluster Bitcoin addresses according to two heuristics: one exploits an inherent property of the Bitcoin protocol, and another exploits an idiom of use in the Bitcoin network. By layering this clustering analysis on top of our ground-truth data collection, we transitively taint entire clusters of addresses as belonging to certain users and services; e.g. if our analysis indicated that an address we had previously tagged as belonging to Bitstamp was contained in a certain cluster, we could tag – with some reasonable level of confidence – all of the addresses in that cluster as belonging to Bitstamp as well.

As of 13 April, 2013, when we conducted this research, the blockchain contained over 16 million transactions carried out between 12 million distinct addresses. Over 11 million bitcoins had been generated (recall this is over half of all the bitcoins that will ever be generated) and those bitcoins had been spent many times over, to the point that over 1 trillion bitcoins had been transacted.

A re-identification attack

The first phase of our analysis involved interacting with many of the different types of services described earlier in order to perform our ground-truth data collection. In total, we kept accounts with 26 exchanges and ten wallet services, and made purchases with 25 different vendors, nine of which used the payment gateway BitPay. We also participated in Bitcoin's mining economy, which involved joining a 'pool' of miners who use their collective resources to mine blocks (as doing so individually requires a significant investment of computational resources); this meant purchasing an AMD Radeon HD 7970 graphics card and engaging with 11 different mining pools. We also kept accounts with five poker sites, and transacted with eight sites offering mini-games and/or lotteries. Finally, we sent bitcoins through four so-called 'mix' services, who promise to send different bitcoins from the ones that are sent to them (thus voiding the transaction history and essentially cleaning any potentially dirty bitcoins), and interacted with a handful of additional miscellaneous sites, including two donations to Wikileaks.

We engaged in 344 transactions with these services, which allowed us to definitively tag 832 addresses (recall that transactions can have arbitrarily many input addresses, which allows us to tag multiple addresses per transaction). We additionally scraped various publicly claimed addresses that we found, such as users' signatures in Bitcoin forums, although we were careful to use only tags for which we could perform some manual due diligence.

Clustering Bitcoin addresses

In theory, the use of pseudonyms within Bitcoin provides a property called *unlinkability*, which says that users' transactions using one set of pseudonyms should not be linked to their

transactions using a different set of pseudonyms. In practice, however, certain properties of Bitcoin usage erode this anonymity.

Recall that, in order to create a valid Bitcoin transaction, the sender must know the private signing key corresponding to the public key in which the bitcoins are held. Now suppose that a user wishes to send 10 BTC to a merchant, but has 4 BTC in one address and 6 BTC in another. One potential way to pay the merchant would be to create a new address, send the 4 BTC and 6 BTC to this new address, and then send the 10 BTC now contained in this new address to the merchant. (In fact, this is the method that preserves the most anonymity.) Instead, the Bitcoin protocol allows for a simpler and more efficient solution: transactions can have arbitrarily many inputs, so the 4 BTC and 6 BTC addresses can be used as inputs to the same transaction, in which the receiver is the merchant.

This observation gives rise to our first clustering heuristic: if two addresses have been used as input to the same transaction, they are controlled by the same user. This heuristic is quite safe (or at least it was in April 2013), as the sender must know the private keys corresponding to all input addresses in order to form a valid transaction, and as such it has already been used in the Bitcoin literature and free tools are available online that perform this analysis.

Our second clustering heuristic expands on this first heuristic and exploits the way in which change is made. In the Bitcoin protocol, when an address receives some number of bitcoins, it has no choice but to spend those bitcoins all at once (recall that this is because each transaction must reference a previous transaction, and transactions cannot be referenced multiple times). If this number of bitcoins is in excess of what the sender wants to spend (e.g. if he has 4 BTC stored in an address and wants to send 3 BTC to a merchant), then he creates a transaction with two outputs: one for the actual recipient (e.g. the merchant receiving 3 BTC) and one *change address* that he controls and can use to receive the change (e.g. the 1 BTC left over).

This behaviour gives rise to our second clustering heuristic: the change address in a transaction is controlled by the sender. As change addresses do not a priori look any different from other addresses, significant care must be taken in identifying them. As a first step, we observe that in the standard Bitcoin client, a change address is created internally and is not even known to the user (although he can always learn it by examining the blockchain manually). Furthermore, these change addresses are used only twice: once to receive the change in a transaction, and once to fully spend their contents as the input in another transaction (in which the client will create a fresh address to receive any change).

By examining transactions and identifying the outputs that meet this pattern of one-time usage, we identify the change addresses (if more than one output meets this pattern, then we err on the side of safety and do not tag anything as a change address). Using this pattern – with a number of additional precautions, such as waiting a week to identify change addresses – we identified 3.5 million change addresses, with an estimated false positive rate of 0.17 per cent (note that the false positive rate can only be estimated, as in the absence of ground-truth data we cannot know what truly is and isn't a change address). By then clustering addresses according to this heuristic, we collapsed the 12 million public keys into 3.3 million clusters.

Putting it together

By layering our clustering analysis on top of our ground-truth data (and thus transitively tagging entire clusters that contain previously tagged addresses), we were able to identify 1.9 million public keys with some real-world service or identity, although in many cases the identity was not a real name, but rather (for example) a username on a forum. While this is a somewhat

small fraction (about 16 per cent) of all public keys, it nevertheless allows us to de-anonymise significant flows of bitcoins throughout the network.

Towards this goal, we first examined interactions with known Bitcoin services. By identifying a large number of addresses for various services (e.g. we identified 500,000 addresses as controlled by Mt. Gox, the biggest exchange at the time of our analysis, and over 250,000 addresses as controlled by the underground marketplace Silk Road), we were able to observe interactions with these services, such as deposits into and withdrawals from exchanges. While this does not de-anonymise the individual participating in the transaction (i.e. we could see that a user was interacting with a service, but did not necessarily know which user), it does serve to de-anonymise the flow of bitcoins into and out of the service.

Tracking flows of Bitcoins

To demonstrate the usefulness of this type of analysis, we turned our attention to criminal activity. In the Bitcoin economy, criminal activity can appear in a number of forms, such as dealing drugs on Silk Road or simply stealing someone else's bitcoins. In this study, we followed the flow of bitcoins out of Silk Road (in particular, from one notorious address) and from a number of highly publicised thefts to see if we could track the bitcoins to known services. While some of the thieves attempted to use sophisticated mixing techniques (or possibly mix services) to obscure the flow of bitcoins, for the most part tracking the bitcoins was quite straightforward, and we ultimately saw large quantities of bitcoins flow to a variety of exchanges directly from the point of theft (or the withdrawal from Silk Road).

Our tracking technique focused on one particular idiom of use in Bitcoin that we call a 'peeling chain.' A peeling chain begins with a single address that holds a relatively large amount of bitcoins; as a concrete example, imagine that an exchange has 100 BTC stored in a single address. A smaller amount is then 'peeled' off this larger amount, creating a transaction in which a small amount is transferred to one address and the remainder is transferred to a change address; continuing with the example, this could represent a user withdrawing 1 BTC from an account, in which case the bank would peel off 1 BTC and send the remaining 99 BTC to a change address that it controls. This process is repeated for potentially many 'hops' until the larger amount is pared down to a small fraction of a bitcoin, at which point the amount remaining in the address could be aggregated with others to begin the peeling process all over again. We can repurpose our clustering heuristics to follow these peeling chains, as our ability to identify change addresses means we can isolate the 'peel' (i.e. the meaningful recipient in a transaction) and continue to follow the chain using change addresses.

Our reason for targeting exchanges was twofold. First, these are public-facing services that interact with fiat currency and its surrounding regulatory landscape; as such, they are increasingly (and necessarily) starting to implement know-your-customer (KYC) policies and are thus the likeliest service to know the real-world identity of Bitcoin users. Second, if one wants to cash out of the Bitcoin economy it is (or at least was in 2013) quite difficult to do so without using an exchange: they serve as choke points into and out of the system. Thus, while tracking stolen bitcoins to their point of deposit into an exchange does not in itself identify the thief, our analysis could be coupled with subpoena power to potentially identify the real-world owner of the account into which the stolen bitcoins were deposited.

Tracking bitcoins out of Silk Road

We first applied our tracking technique to an address believed to be associated with Silk Road. This address is one of the most well-known Bitcoin addresses: at its peak it contained 5 per cent

of all generated bitcoins, receiving 613,326 BTC over a period of only eight months and then promptly emptying its contents almost immediately.

In emptying its contents, this Silk Road address sent 158,336 bitcoins to another address, which in turn sent 50,000 to each of two addresses and 58,336 to another; each of these three addresses then began a peeling chain consisting of hundreds of hops. We followed each of these chains for 100 hops and observed peels to ten different Bitcoin exchanges (including, in the first chain, 11 peels to Mt. Gox that collectively sent 492 BTC to the exchange); out of the 300 peels along the chains, 54 of them went to these exchanges. As described above, this provides an opportunity to find the real-world identity associated with this user for anyone able to subpoena these exchanges.

Tracking thefts

To ensure that our analysis could be applied more generally, we also looked at some of the major Bitcoin heists that had taken place in its history, ranging from the theft of 58,547 BTC from Bitcoinica, the largest margin-trading service and one of the most popular services in Bitcoin's history, to bitcoin-stealing malware that succeeded in stealing 3,257 BTC from various individual victims.

In examining these thefts, we observed that certain thieves attempted relatively sophisticated patterns of layering and mixing, while others made almost no attempt to hide the source of the stolen bitcoins. This setting thus provided a useful demonstration of the potential for anonymity provided by Bitcoin, and the ways in which certain usage would cause one to fall short of this potential.

Conclusions and implications for crime prevention

This study attempted to provide a characterisation of the evolving nature of Bitcoin, focusing on the rise of public-facing services and the growing gap between the potential anonymity available in the Bitcoin protocol and the actual anonymity that is achieved by its users. To this end, we developed new clustering heuristics for identifying Bitcoin addresses that belong to the same user, and used manual data collection to hand-label a small set of Bitcoin transactions; combining these two approaches allowed us to associate 1.9 million Bitcoin addresses with named individuals or services.

As acknowledged above, these techniques and the tracking technique they enable do not in themselves serve to de-anonymise individual users of Bitcoin. Rather, we argue that they enable further de-anonymisation in the cases in which certain agencies can determine – through, for example, subpoena power – the real identities of users interacting with services such as exchanges.

Thus, while this work does not in itself prevent anyone from using Bitcoin for criminal activity, the ease with which our analysis could be applied led us to conclude that using Bitcoin for money laundering or other illicit purposes did not (at least at the time of our study) seem particularly attractive. Based on the informally observed activity of thieves subsequent to our study, this deterrence seems to have served as a form of prevention, or at least as a barrier to entry, as now only those who believe themselves expert enough in the methods for achieving anonymity can safely use Bitcoin for criminal activities.

Note

1 In here and what follows, we use Bitcoin to mean the peer-to-peer network and abstract protocol, and bitcoin, or BTC, to mean the unit of currency.